

Appendix

Appendix A1 Study characteristics: Baker, 1997 (randomized controlled trial)

Characteristic	Description
Study citation	Baker, J. J. (1997). Effects of a generative instructional design strategy on learning mathematics and on attitudes towards achievement. <i>Dissertation Abstracts International</i> , 58(7), 2573A. (UMI No, 9800955).
Participants	Ninety eighth-grade students. Most students were from low-income families and qualified for free or reduced-price lunches. All but three students were white. None were in special education. Students were randomized to the intervention or the comparison condition. Seventy students completed the math pretest and posttest.
Setting	Suburban middle school in St. Louis, Missouri; four classrooms (two intervention classrooms and two comparison classrooms).
Intervention	The intervention group was taught using a “generative mathematics curriculum” that used <i>The Expert Mathematician</i> (version 3.0). Students worked individually or in pairs using the printed materials and the computer to work through the lessons in <i>The Expert Mathematician</i> . Sessions, which included one or two lessons, were 85 minutes long and occurred every other day for one school year. Intervention students were taught in separate classrooms from comparison students, but the same teachers taught both groups.
Comparison	The comparison group experienced a “linear mathematics curriculum” based on <i>Transition Mathematics</i> , the middle school volume of the University of Chicago School Mathematics Project. The author describes this as a traditional, teacher-directed curriculum. The text covers the first year in a six-year mathematics curriculum.
Primary outcomes and measurement	The 78-item Objectives by Strand test was developed by the district. No norming information was available. The test was administered at the end of the school year by the classroom teacher. (See Appendix A2 for more detailed descriptions of the outcome measure.)
Teacher training	None reported.

Appendix A2 Outcome measure in the math achievement domain

Outcome measure	Description
Objectives by Strand	A 78-item multiple choice test of mathematical ability developed by a large urban school district in 1980 and administered at the end of the school year. Includes 61 concepts and applications items. Average internal consistency (Cronbach's alpha) across the intervention and comparison groups was greater than 0.90. No information on norms was available.

Appendix A3 Summary of study findings included in the rating for the math achievement domain¹

			Author’s findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
					Mean difference ⁴ (<i>Expert Mathematician</i> – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
Outcome measure	Study sample	Sample size (schools/ students)	<i>Expert Mathema- tician</i> group ³	Comparison group				
Baker, 1997 (randomized controlled trial) ⁸								
Objectives by Strand test	8th graders	70 students/ 4 classrooms	45.10 (12.03)	40.80 (12.41)	4.30	0.35	ns	+14
Domain average ⁹ for math achievement						0.35	ns	+14

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. In the case of Baker (1997), a correction for pretest differences was applied to posttest results.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, please see [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Baker (1997), no corrections were needed.
9. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 *The Expert Mathematician* rating for the math achievement domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of math achievement, the WWC rated *The Expert Mathematician* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *The Expert Mathematician* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. The one study reviewed by the WWC showed a substantively important positive effect.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. The WWC analysis found no studies with statistically significant or substantively important negative effects or with indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. There is only one study in this domain, which did not show statistically significant positive effects.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. The one study reviewed by the WWC did not show statistically significant negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.